20000814108

(2)

MRC Technical Summary Report #2825

# A LINEAR COMBINATION TEST FOR DETECTING SERIAL CORRELATION IN MULTIVARIATE SAMPLES


Richard A. Johnson and Thore Langeland





**Mathematics Research Center**
**University of Wisconsin—Madison**
**610 Walnut Street**
**Madison, Wisconsin 53705**





June 1985

(Received March 27, 1985)







Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, D. C. 20550

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

# A LINEAR COMBINATION TEST FOR DETECTING SERIAL CORRELATION IN MULTIVARIATE SAMPLES

Richard A. Johnson and Thore Langeland

Technical Summary Report #2825
June 1985

## ABSTRACT

~~We~~ propose a test for detecting serial dependence among multivariate observations. ~~Our~~ The test statistic is the maximum absolute value of the lag 1 correlation obtainable from a linear combination of the observations. ~~We~~ The authors express the statistic in terms of two eigenvalues and then obtain the asymptotic null distribution. Asymptotic power is examined for sequences of local alternatives in a multivariate normal autoregressive process. An explicit expression is obtained for the density of the limit distribution in the bivariate case. ~~We~~ They then compare power with the likelihood ratio statistic.

AMS (MOS) Subject Classifications: 62H10, 62E20

Key Words: Multivariate, Test of independence, Linear combination, Distribution of eigenvalues

Work Unit Number 4 (Statistics and Probability)

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the National Science Foundation under Grant No. MCS-820-2205.

## SIGNIFICANCE AND EXPLANATION

If multivariate observations taken at adjacent times are correlated the quality of inferences, based on an independence assumption, can be seriously eroded. After illustrating these effects, we propose a new test for detecting dependence among adjacent observations. We reduce the problem to one dimension by considering linear combinations $a_1X_1 + \cdots + a_kX_k$ of the observations.

Our test statistic is then the maximum, over all linear combinations, of the sample auto-correlation. We determine its large sample null distribution from which approximate critical values can be obtained numerically.

Because of the seriousness of departures from independence, it is important to have procedures for detecting dependence. Our statistic provides one alternative way of quantifying dependence in a series of multivariate observations. It should be a useful addition to summary descriptions of multivariate data sets and serve as a warning when multivariate time series methods are required.



The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

A LINEAR COMBINATION TEST FOR DETECTING SERIAL CORRELATION IN MULTIVARIATE SAMPLES

Richard A. Johnson and Thore Langeland

## 1. INTRODUCTION

It is well-known that the presence of even a moderate autocorrelation, among univariate observations, can cause serious difficulties for procedures based on an assumption of independence. In the multivariate case, both inferences about the mean $\underset{\sim}{\mu}$, and covariance matrix, $\Sigma$, can be severely affected. To illustrate, let $\underset{\sim}{X}_t$ follow the multivariate AR(1) model

$$\underset{\sim}{X}_t - \underset{\sim}{\mu} = \Phi(\underset{\sim}{X}_{t-1} - \underset{\sim}{\mu}) + \underset{\sim}{\varepsilon}_t$$

where the $\underset{\sim}{\varepsilon}_t$ and independent and identically distributed with $E(\underset{\sim}{\varepsilon}_t) = \underset{\sim}{0}$ and $\mathrm{Cov}(\underset{\sim}{\varepsilon}_t) = \underset{\sim}{\Sigma}_\varepsilon$ and all the eigenvalues of $\Phi$ are between $-1$ and $1$. As a consequence of the ergodic theorem

$$\bar{\underset{\sim}{X}} \xrightarrow{a.s.} \underset{\sim}{\mu}, \quad S = \frac{1}{n-1}\sum_{t=1}^{n}(\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})' \xrightarrow{a.s.} \underset{\sim}{\Sigma}_X .$$

Also

$$\mathrm{Cov}(n^{-1/2}\sum_{t=1}^{n}\underset{\sim}{X}_t) \to (I - \Phi)^{-1}\underset{\sim}{\Sigma}_X + \underset{\sim}{\Sigma}_X(I - \Phi')^{-1} - \underset{\sim}{\Sigma}_X$$

and $\sqrt{n}(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})$ is asymptotically normal with this limiting covariance matrix. If the underlying process has $\Phi = \phi I_k$, $|\phi| < 1$, then the nominal 95% confidence ellipsoid $\{\underset{\sim}{\mu} : n(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu})'S^{-1}(\bar{\underset{\sim}{X}} - \underset{\sim}{\mu}) \le \chi_k^2(.05)\}$ has limiting coverage probability $P[\chi_k^2 \le \frac{1-\phi}{1+\phi}\chi_k^2(.05)]$. For instance if $\underset{\sim}{X}_t$ has dimension $k = 5$ and $\phi = .3$ the coverage probability is .690. For $k = 10$ and $\phi = .5$ it is .193.

---

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the National Science Foundation under Grant No. MCS-820-2205.

In the context of principal component analysis, suppose we wish to analyze $\Sigma_{\varepsilon}$, which is the covariance matrix for $X_t$ under independence, but that the correlation structure is introduced by selecting a sampling interval that is too short. The first principal component has coefficient vector $e_1$ where $\Sigma_{\varepsilon} e_1 = \lambda_1 e_1$, with $\lambda_1 > \cdots > \lambda_p$. However, if the underlying process has $\Phi = C \Sigma_{\varepsilon}^{-1}$ where $C$ is just smaller than $\lambda_p$,

$$\Sigma_X e = \frac{\lambda^3}{\lambda^2 - C^2} e$$

so $e_p$ is incorrectly identified as the coefficient vector of the first principal component.

Numerous tests have been proposed for the univariate case. Ligget (1977), Bartlett and Rajalaksham (1953), and Chittu...i (1974) have proposed multivariate extensions of the Bartlett periodogram test, the Quenouille test and Box and Pierce test, respectively.

## 2. A LINEAR COMBINATION TEST

Because first order autocorrelation is most common, it is worthwhile to develop a test for first order correlation that is both easy to apply and has a graphic interpretation. We reduce the problem to one dimension by considering linear combinations $\underset{\sim}{a}'\underset{\sim}{X}_t$, $t = 1,2,\ldots,T$ and selecting $\underset{\sim}{a}$ to maximize the lag 1 correlation

$$r_{\underset{\sim}{a}}(1) = \frac{\sum_{t=1}^{T-1} \underset{\sim}{a}'(\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_{t+1} - \bar{\underset{\sim}{X}})'\underset{\sim}{a}}{\sum_{t=1}^{T} [\underset{\sim}{a}'(\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})]^2} = \frac{\underset{\sim}{a}'C_1\underset{\sim}{a}}{\underset{\sim}{a}'C_0\underset{\sim}{a}}$$

where the sample cross-covariance matrix of lag $j$ is

$$C_j = \frac{1}{T}\sum_{t=1}^{T-j} (\underset{\sim}{X}_t - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_{t+j} - \bar{\underset{\sim}{X}})' \quad \text{for} \quad j = 0,1,\ldots,T-1\,. \tag{2.1}$$

Our test statistic is then defined as the maximum attainable lag 1 correlation,

$$R_L = \sup_{\underset{\sim}{a}\neq\underset{\sim}{0}} |r_{\underset{\sim}{a}}(1)|\,.$$

Setting $C_s = 2^{-1}(C_1 + C_1')$, $r_{\underset{\sim}{a}}(1)$ can be expressed in terms of symmetric matrices as

$$R_L = \sup_{\underset{\sim}{a}\neq\underset{\sim}{0}} \frac{|\underset{\sim}{a}'C_s\underset{\sim}{a}|}{\underset{\sim}{a}'C_0\underset{\sim}{a}} = \max\{|\hat{\lambda}_1|, \hat{\lambda}_k\} \tag{2.2}$$

where $\hat{\lambda}_1 \le \hat{\lambda}_2 \le \cdots \le \hat{\lambda}_k$ are the eigenvalues of $C_0^{-1/2}C_sC_0^{-1/2}$ or $C_0^{-1}C_s$. One point of difficulty is that $C_s$ is not necessarily non-negative definite.

Note that $R_L$ has the properties

(i) $R_L \ge |r_i(1)|$, $r_i = \dfrac{\sum_{t=1}^{T-1} (x_{ti} - \bar{x}_i)(x_{t+1,i} - \bar{x}_i)}{\sum_{t=1}^{T} (x_{ti} - \bar{x}_i)^2}\,.$

(ii) $R_L$ is invariant under

$$X_t \to A X_t Q$$

where $A$ is non-singular and $Q$ orthogonal.

A plot of $\hat{a}'(X_t - \bar{X})$ versus $\hat{a}'(X_{t+1} - \bar{X})$ displays the concentrated correlation estimated by $R_L$.

We now indicate the steps leading to the asymptotic null distribution for $R_L$ leaving the more technical algebraic steps until Section 5. We say that the $k \times k$ matrix $B$ is $N_{k^2}(0, \Sigma \otimes \Sigma^{-1})$ if $\mathrm{tr}(A'B)$ is $N(0, \mathrm{tr} A\Sigma A'\Sigma^{-1})$ for every $k \times k$ matrix $A$. Mann and Wald (1943) showed that

$$T^{1/2} C_0^{-1} C_1 \xrightarrow{\mathcal{L}} N_{k^2}(0, \Sigma \otimes \Sigma^{-1})$$

so $T^{1/2} C_0^{-1/2} C_s C_0^{-1/2} \xrightarrow{\mathcal{L}} S$ where, under the null hypothesis, $S$ has pdf

$$f(S) = (2\pi)^{-k(k+1)/4} 2^{k(k-1)/4} \mathrm{etr}(-\tfrac{1}{2} SS) , \qquad (2.3)$$

with respect to $k(k+1)/2$ dimensional Lebesgue measure.

Hsu (1939) encountered the same asymptotic distribution while studying a normal theory one-way MANOVA problem. He established that, if $S$ is distributed as (2.3), the distribution of its eigenvalues $\lambda_1 < \cdots < \lambda_k$ has pdf

$$g(\lambda_1, \lambda_2, \ldots, \lambda_k) = [2^{k/2} \prod_{i=1}^{k} \Gamma(i/2)]^{-1} \prod_{i<j}^{k} (\lambda_j - \lambda_i) \cdot e^{-\sum_{i=1}^{k} \lambda_i^2/2} . \qquad (2.4)$$

Since $T^{1/2} R_L$ is a continuous function of $T^{1/2} C_0^{-1/2} C_s C_0^{-1/2}$,

$$\sqrt{T}\, R_L \xrightarrow{\mathcal{L}} \max(|\lambda_1|, \hat{\lambda}_k) . \qquad (2.5)$$

For $k = 2$, the limit distribution is easy to evaluate

$$P[T^{1/2} R_L \le x] \to P[-x < \Lambda_1 < \Lambda_2 < x] = \sqrt{2} \int_{-x}^{x} u e^{-u^2/2} \Phi(u) du = F(x) . \qquad (2.6)$$

It is considerably more difficult to present expressions for the general case. Set

$$G_j(t) = \int_{-x}^{t} u^j e^{-u^2/2} du, \quad j = 0,1,2,\ldots,k, \tag{2.7}$$

$$G_{j,\ell}(x) = \int_{-x}^{x} G_j(t) t^{\ell} e^{-t^2/2} dt, \quad 0 \le j,\ell \le k \tag{2.8}$$

where it can be shown (see Mehta (1960), p. 399, eqn. (13))

$$G_{j,\ell}(x) = (-1)^{\ell+j} G_{\ell,j}(x) . \tag{2.9}$$

In Section 5, we establish

Theorem 2.1. For $k$ even, the asymptotic cdf of the LCT statistic $T^{1/2}R_L$, under the null hypothesis of independence, is

$$F(x) = \left(\prod_{i=1}^{k} \Gamma(i/2)\right)^{-1} \det(\{G_{j,\ell}(x)\})$$

for $j = 0,2,4,\ldots,k-2$ and $\ell = 1,3,5,\ldots,k-1$, where $G_{j,\ell}(x)$ is defined in (2.8).

Theorem 2.2. For $k$ odd, the asymptotic cdf of the LCT statistic $T^{1/2}R_L$, under the null hypothesis of independence, is

$$F(x) = \left[2^{1/2} \prod_{i=1}^{k} \Gamma(i/2)\right]^{-1} \sum_{j=0}^{(k-1)/2} (-1)^{(k-1)/2+j} G_{2j}(x) \det(B_j)$$

where $G_j(x)$ is defined in (2.7),

$$B_j = \begin{bmatrix} G_{0,1}(x) & G_{0,3}(x) & \cdots & G_{0,k-2}(x) \\ G_{2,1}(x) & G_{2,3}(x) & \cdots & G_{2,k-2}(x) \\ \vdots & \vdots & & \vdots \\ G_{2j-2,1}(x) & G_{2j-2,3}(x) & \cdots & G_{2j-2,k-2}(x) \\ G_{2j+2,1}(x) & G_{2j+2,3}(x) & \cdots & G_{2j+2,k-2}(x) \\ \vdots & \vdots & & \vdots \\ G_{k-1,1}(x) & G_{k-1,3}(x) & \cdots & G_{k-1,k-2}(x) \end{bmatrix}$$

for $j = 0,1,2,\ldots,(k-1)/2$, and $G_{j,\ell}(x)$ is defined in (2.8).

A table of 1-st, 5-th and 10-th percentiles, for $k = 2(1)20$ were calculated using double precision arithmetic (see Langeland (1980)).

## 3. SC[illegible] COMPETING TESTS AND POWER CONSIDERATIONS

Most tests for independence are motivated from consideration of autoregressive alternatives. Let

$$\underset{\sim}{X}_t - \underset{\sim}{\mu} = \Phi(\underset{\sim}{X}_{t-1} - \underset{\sim}{\mu}) + \underset{\sim}{\varepsilon}_t$$

for $t = 1,2,\ldots,T$. The hypothesis of independence is then

$$H : \Phi_1 = 0 . \tag{3.1}$$

A natural test statistic to use is

$$S_L = -[N - \frac{1}{2}(k + k + 1)]\log(|C_0 - \hat{\Phi}C_0\hat{\Phi}'|/|C_0|) \tag{3.2}$$

where $N = T - 1 - 1$ and $\hat{\Phi} = C_1C_0^{-1}$. If the $\{\underset{\sim}{\varepsilon}_t\}_{t=1}^{T}$ are i.i.d. multivariate normal, then the test statistic in (3.2) has the same asymptotic distribution as the logarithm of the likelihood ratio test statistic. See Hannan (1970, pages 338-341).

Theorem 3.1. Under the null hypothesis of independence (3.1), the asymptotic distribution of the test statistic (3.2) is a $\chi^2_{k^2}$-distribution.

In order to obtain an indication of asymptotic power, we introduce the normal theory AR(1) model (3.1) where the $\underset{\sim}{\varepsilon}_t$ are independent $N(\underset{\sim}{0},\underset{\sim}{\Sigma}_\varepsilon)$. Let $\{\Phi_T\}$ be a sequence of alternatives to independence, where $T^{1/2}\Phi_T \to H$, and let $P_{T,\Phi_T}$ denote the distribution of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_T$. Let $P_T$ be the distribution of $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_T$ under independence.

Theorem 3.2. Under $\{P_T\}$

$$\Lambda_T = \ln\frac{dP_{T,\Phi_T}}{dP_T} = \mathrm{tr}[\Sigma_\varepsilon^{-1}T^{1/2}\Phi_T T^{1/2}C_1] - \frac{1}{2}\mathrm{tr}[\underset{\sim}{\Sigma}_\varepsilon T^{1/2}\Phi_T C_0 T^{1/2}\Phi_T'] + o_{P_n}(1)$$

$$\to N(-\frac{1}{2}\sigma^2,\sigma^2)$$

so $\{P_T\}$ and $\{P_{T,\Phi_T}\}$ are contiguous.

It can then be shown that $(\Lambda_T, T^{1/2}C_0^{-1/2}C_sC_0^{-1/2})$ is asymptotically normal under $P_T$ so that we can obtain the limiting distribution of the linear combination statistic, $R_L$, under $\{P_{T,\Phi_T}\}$. Even the bivariate case is complicated. The limit distribution for $T^{1/2}R_L$ is

$$f(x) = 4e^{-(\lambda+\eta)/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{(j!)^2} \sum_{i=0}^{\infty} \frac{(\eta/2)^i}{i!} \frac{x^{2(j+i+1)}}{\Gamma(\frac{2i+1}{2})}$$

$$\cdot \int_0^1 (1-u)^{2j+1} u^{2i} e^{-x^2[u^2+(1-u)^2]} du$$

for $x > 0$, where $\eta = (\mu_1 + \mu_3)^2/2$, $\lambda = [(\mu_1 - \mu_3)^2 + 4\mu_2^2]/2$ and

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} (\underset{\sim}{\Phi}_\varepsilon^{1/2} \otimes \underset{\sim}{\Phi}_\varepsilon^{-1/2}) \mathrm{vec}(\Phi) .$$

It also follows directly that $(\Lambda_T, S_L)$ are each jointly normal under $\{P_T\}$. From the continguity, we then obtain

<u>Theorem 3.3</u>. Under $\{P_{T,\phi_T}\}$, the asymptotic distribution of $S_L$ is non-central $\chi^2_{k^2}$ with noncentrality parameter $tr[\underset{\sim}{\Phi}_\varepsilon^{-1} H \Sigma_{\underset{\sim}{\varepsilon}} H']$.

It is well-known that the likelihood ratio test has several large sample optimal properties. However, a calculation of asymptotic power in Table 3.1 with $k = 2$, $\underset{\sim}{\Phi}_\varepsilon = I$ shows that the linear combination test has higher power than the others when $T^{1/2}\phi_T \to diag(h_{11}, 0)$.

Table 3.1

Asymptotic Power

| $h_{11}$ | $R_L$ | $S_L$ |
|---|---|---|
| .1 | .0513 | .0505 |
| .5 | .0849 | .0627 |
| 1 | .1796 | .1055 |
| 2 | .4666 | .3201 |
| 3 | .7714 | .6635 |
| 5 | .9952 | .9894 |

## 4. EXAMPLE

We consider some data reported by Simon (see Duncan (1959), pages 626-630) consisting of burning times of 30 fuses as recorded by three observers. Since there is one missing observation for the second observer, we first confine ourselves to the data given by observers one and three. Let $\underset{\sim}{X}_t = (X_{1,t}, X_{2,t})'$, $t = 1,2,\ldots,30$ denote the observations. The plot of $X_{i,t}$ versus $X_{i,t+1}$ for $i = 1$ is given below in Figure 4.1. The plot for $i = 2$ is similar. Neither exhibits clear signs of first order serial dependence. The LCT statistic $\sqrt{30}\, R_L = 2.40$ and it is significant at the 10 percent level. The value of the corresponding eigenvector is $\hat{\underset{\sim}{a}} = (1.0,-.99)'$. The plot of $\hat{a}'X_t$ versus $\hat{a}'X_{t+1}$ given in Figure 4.2 gives an indication of serial dependence in the two series of data. If the missing observation is estimated, the evidence for dependence with three observers is much stronger. The statistic becomes significant at the 3% level.

Figure 4.1

PLOT OF DATA OF OBSERVER ONE VERSUS THESE DATA LAGGED ONE UNIT

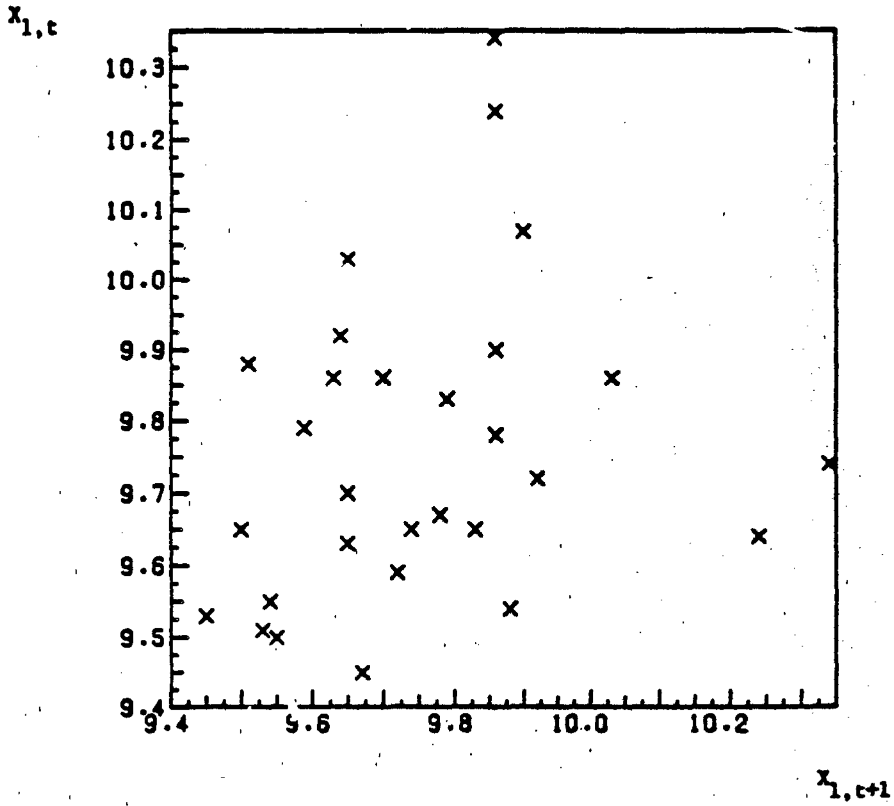
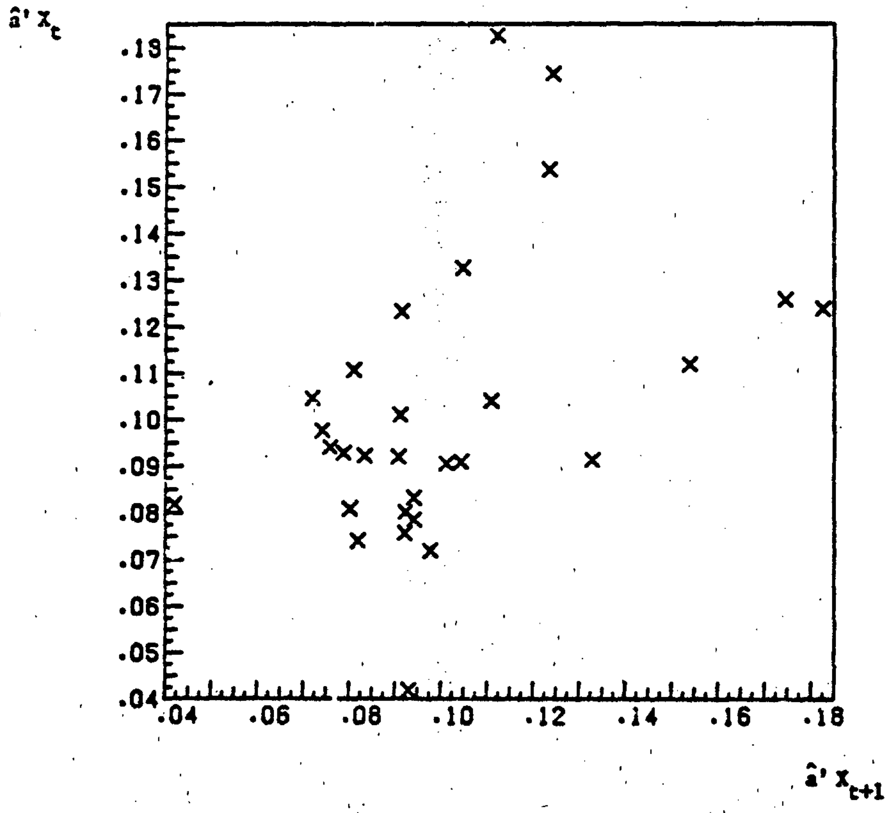


Figure 4.2

PLOT OF $\hat{a}'X_t$ VERSUS $\hat{a}'X_{t+1}$ FOR DATA FOR OBSERVERS ONE AND THREE



## 5. DERIVATIION OF LIMITING NULL DISTRIBUTION

The asymptotic cdf of $T^{1/2}R_L$ is given by

$$F(x) = P[-x < \Lambda_1 < \Lambda_k < x] = \int \cdots \int_{Q(-x,x)} g(\lambda_1,\ldots,\lambda_k)d\lambda_1 \cdots \lambda_k \tag{5.1}$$

where $g(\cdot)$ is defined in (2.4) and $Q(a,b) = \{a < \lambda_1 < \lambda_2 \cdots < \lambda_k < b\}$. Since

$$\prod_{1\le i<j\le k} (\lambda_j - \lambda_i) = \det\left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_k \\ \vdots & \vdots & & \vdots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \cdots & \lambda_k^{k-1} \end{bmatrix}\right)$$

(the Vandermonde determinant), (5.1) can be rewritten as

$$\int \cdots \int_{Q(-x,x)} c_k \det\left(\begin{bmatrix} e^{-\lambda_1^2/2} & e^{-\lambda_2^2/2} & \cdots & e^{-\lambda_k^2/2} \\ \lambda_1 e^{-\lambda_1^2/2} & \lambda_2 e^{-\lambda_2^2/2} & \cdots & \lambda_k e^{-\lambda_k^2/2} \\ \vdots & \vdots & & \vdots \\ \lambda_1^{k-1} e^{-\lambda_1^2/2} & \lambda_2^{k-1} e^{-\lambda_2^2/2} & \cdots & \lambda_k^{k-1} e^{-\lambda_k^2/2} \end{bmatrix}\right) \cdot d\lambda_1 \cdot d\lambda_2 \cdots d\lambda_k \tag{5.2}$$

where $c_k = [2^{k/2} \sum_{j=1}^{k} \Gamma(j/2)]^{-1}$.

In order to obtain an explicit expression for the densities we need some additional concepts and lemmas (see Aitken (1939), pages 50 and 111).

The signature function $E(x_1,x_2,\ldots,x_k)$ is defined as

$$E(x_1,x_2,\ldots,x_k) = \prod_{1\le i<j\le k} \operatorname{sign}(x_j - x_i) \tag{5.3}$$

for $x = (x_1,x_2,\ldots,x_k)' \in R^k$, $E(x_1,x_2,\ldots,x_k) = 0$ if $x_i = x_j$ for some $i \ne j$, $i,j = 1,2,\ldots,k$, and $E(x_1) = 1$ for all $x_1 \in R$.

Let $k = 2m$ and $m = 1,2,\ldots,$ and let $A = \{a_{ij}\}$ be a skew $(k \times k)$ matrix, then the Pfaffian of $A$, $\mathrm{Pf}(A)$, is defined as

$$\mathrm{Pf}(A) = (2^m m!)^{-1} \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} \cdots \sum_{j_k=1}^{k} E(j_1,j_2,\ldots,j_k)$$

$$\cdot\, a_{j_1 j_2} \cdot a_{j_3 j_4} \cdots a_{j_{k-1} j_k} \,.$$

It is well-known that $[\mathrm{Pf}(A)]^2 = \det A$.

de Bruijn (1955) has established the following expression for $k$ even.

Lemma 5.1. Assume $\det(\{\phi_j(x_j)\}) \in L(\mathbb{R}^k)$ and let $k = 2m$ and $m = 1,2,\ldots,$ then

$$\int_{Q(a,b)} \cdots \int \det(\{\phi_j(x_i)\})dx_1 dx_2 \cdots dx_k$$

$$= \mathrm{Pf}(\{a_{ij} = \int_a^b \int_a^b \phi_i(x)\phi_j(y)\,\mathrm{sign}(y-x)\,dxdy\}) \,. \qquad (5.4)$$

Remark. de Bruijn (1955) gives a somewhat unusual definition of the Pfaffian and his derivation of the integral on the left-hand side of (5.4), for $k$ odd, is only valid in a very special case. However, Krishnaiah and Chang (1971, equation 2.6) give a general solution to the odd case. In their notation $\phi_j(x) = x^{r+j-1}\psi(x)$ for $r > 0$ and some function $\psi(x)$ satisfying the integrability conditions. We restate their results as Lemma 5.2 (an alternative proof is given in Langeland (1980)).

Lemma 5.2. Assume $\det(\{\phi_j(x_i)\}) \in L(\mathbb{R}^k)$ and let $k$ be odd, then

$$\int_{Q(a,b)} \cdots \int \det(\{\phi_j(x_i)\})dx_1 dx_2 \cdots dx_k = \sum_{j=1}^{k} (-1)^{j-1}\psi_j(b)\mathrm{Pf}(A_j) \,,$$

where

$$\psi_j(b) = \int_a^b \phi_j(t)dt \quad \text{for} \quad j = 1,2,\ldots,k \,,$$

and

$$A_j = \begin{bmatrix} 0 & a_{12} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,k} \\ a_{21} & 0 & \cdots & a_{2,j-1} & a_{2,j+1} & \cdots & a_{2,k} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{j-1,1} & a_{j-1,2} & \cdots & 0 & a_{j-1,j+1} & \cdots & a_{j-1,k} \\ a_{j+1,1} & a_{j+1,2} & \cdots & a_{j+1,j-1} & 0 & \cdots & a_{j+1,k} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,j-1} & a_{k,j+1} & \cdots & 0 \end{bmatrix}$$

for $j = 1,2,\ldots,k$, and $Q(a,b)$ and $a_{ij}$ are as in Lemma 5.1.

We can now establish Theorem 2.1.

Proof of Theorem 2.1. First we notice that

$$\begin{aligned} &\int_{-x}^{x} [\int_{-x}^{x} u^j e^{-u^2/2} t^\ell e^{-t^2/2} \operatorname{sign}(t-u)du]dt \\ &= \int_{-x}^{x} t^\ell e^{-t^2/2}[\int_{-x}^{t} u^j e^{-u^2/2}du - \int_{t}^{x} u^j e^{-u^2/2}du]dt \\ &= G_{j,\ell}(x) - \int_{-x}^{x} t^\ell e^{-t^2/2}[\int_{t}^{x} u^j e^{-u^2/2}du]dt \\ &= G_{j,\ell}(x) - \int_{-x}^{x} u^j e^{-u^2/2}[\int_{-x}^{u} t^\ell e^{-t^2/2}dt]du \\ &= G_{j,\ell}(x) - G_{\ell,j}(x) \quad \text{for} \quad 0 < j,\ell < k - \ell . \end{aligned}$$

By (2.9), the last quantity equals 0 or $\pm 2G_{j,\ell}(x)$. Lemma 5.1 then gives

$$F(x) = [2^{k/2} \prod_{j=1}^{k} \Gamma(j/2)]^{-1} \tag{5.5}$$

$$\cdot \, \mathrm{Pf}\left(\begin{bmatrix} 0 & 2G_{0,1}(x) & 0 & \cdots & 2G_{0,k-1}(x) \\ 2G_{1,0}(x) & 0 & 2G_{1,2}(x) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 2G_{k-2,1}(x) & 0 & \cdots & 2G_{k-2,k-1}(x) \\ 2G_{k-1,0}(x) & 0 & 2G_{k-1,2}(x) & \cdots & 0 \end{bmatrix}\right) .$$

Let $k = 2m$, then, according to definition of the Pfaffian and the relation for signature functions

$$E(x_1,x_2,\ldots,x_k) = (2^m m!)^{-1} \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} \cdots \sum_{j_k=1}^{k} E(j_1,j_2,\ldots,j_k)$$

$$\cdot \, E(x_{j_1}x_{j_2}) \cdot E(x_{j_3}x_{j_4}) \cdots E(x_{j_{k-1}}x_{j_k})$$

established in de Bruijn (1955), the Pfaffian in (5.5) can be reduced to

$$2^m \sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \cdots \sum_{j_m=1}^{m} E(j_1,j_2,\ldots,j_m) \cdot$$

$$\cdot \, G_{0,2j_1-1}(x), G_{2,2j_2-1}(x), \ldots, G_{k-2,2j_m-1}(x) .$$

But this is nothing but $2^m$ times the determinant in Theorem 2.1. The pr of is complete.

Proof of Theorem 2.2

$$F(x) = [2^{k/2} \sum_{i=1}^{k} \Gamma(i/2)]^{-1} \sum_{j=0}^{k-1} (-1)^j G_j(x) \mathrm{Pf}(A_j)$$

where $A_j = \{a_{pq}\}$ is a $(k-1) \times (k-1)$ matrix with entries $a_{pq} = G_{p,q} - G_{q,p}$ for $p,q = 0,1,\ldots,j-1,j+1,\ldots,k-1$. Next, by (2.7)

$$G_j(x) = 0$$

for $j$ odd. (It can also be shown that $Pf(A_j) = 0$ for $j$ odd.) According to (2.9), for $j$ even, $A_j$ is

$$A_j = 2^{(k-1)} \left[ \begin{matrix} 0 & G_{0,1}(x) & 0 & \cdots & G_{0,j+1}(x) \\ G_{1,0}(x) & 0 & G_{1,2}(x) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ G_{j-1,0}(x) & 0 & G_{j-1,2}(x) & \cdots & 0 \\ G_{j+1,0}(x) & 0 & G_{j+1,2}(x) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ G_{k-2,0}(x) & 0 & G_{k-2,2}(x) & \cdots & 0 \\ 0 & G_{k-1,1}(x) & 0 & & G_{k-1,j-1}(x) \end{matrix} \right.$$

$$\left. \begin{matrix} G_{0,j+1}(x) & \cdots & G_{0,k-2}(x) & 0 \\ 0 & \cdots & 0 & G_{0,k-1}(x) \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & G_{j-1,k-1}(x) \\ 0 & \cdots & 0 & G_{j+1,k-1}(x) \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & G_{k-2,k-1}(x) \\ G_{k-1,j+1}(x) & \cdots & G_{k-1,k-2}(x) & 0 \end{matrix} \right] .$$

All entries containing $j$ as a first or as a second index vanish, i.e., all $G_{\ell,j}(x)$ and $G_{j,\ell}(x)$ for $\ell = 1,3,\ldots,k-2$ vanish. The remaining number of terms $G_{p,q}(x)$, with $p$ even, is exactly $(k-1)/2$. Thus, the Pfaffian of $A_j$ reduces to

$$2^{(k-1)/2} \sum_{j_1=1}^{m} \sum_{j_2=1}^{m} \cdots \sum_{j_m=1}^{m} E(j_1,j_2,\ldots,j_m) G_{0,2j_1-1}(x) \cdot$$

$$\cdot\, G_{2,2j_2-1}(x) \cdots G_{j-2,2j_{(j-2)/2}-1}(x) \cdot G_{j+2,2j_{(j+2)/2}-1}(x) \cdots G_{k-1,2j_m-1}(x)$$

where $m = (k-1)/2$. Except for a possible sign change this is nothing but $2^{(k-1)/2}$ times the determinant of the matrix $B_{(j/2)}$ appearing in the statement of Theorem 2.2. By inspection, the sign is given by $(-1)^{(k-1)/2+j}$. The proof is complete.

We remark that nuclear physicists (e.g. Mehta (1967), Wigner (1967)) are interested in distributions of the eigenvalues of $S$.

## REFERENCES

Aitken, A. (1939) Determinants and Matrices, Oliver and Boyd, Edinburg.

Bartlett, M. S. and D. V. Rajalakshman (1953) Goodness-of-fit Tests for Simultaneous Autoregressive Series, J. Royal Statist. Soc. B 15. 107-124.

Chitturi, R. V. (1974) Distribution of Residuals Autocorrelations in Multiple Autoregressive Schemes, J. Amer. Statist. Assoc. 69, 928-934.

Duncan, A. (1959) Quality Control and Industrial Statistics, Irwin, Homewood, Ill.

de Bruijn, N. (1955) On some Multiple Integrals Involving Determinants, J. Indian Math. Soc. 19, 133-152.

Hannan, E. J. (1970) Multiple Time Series, John Wiley and Sons, New York.

Hsu, P. L. (1939) On the Distribution of the Roots of Certain Determental Equations, Ann. Eugen. 9, 250-258.

Krishnaiah, P. and T. Chang (1971) On the Exact Distributions of the Extreme Roots of the Wishart and MANOVA Matrix, J. Mult. Analysis 1, 108-117.

Langeland, T. (1980) Tests for Dependence in Multivariate Observations, Ph.D. Thesis, University of Wisconsin.

Ligget, W. S. (1977) A Test for Serial Correlation in Multivariate Data, Annals of Statist. 5, 408-413.

Mann, H. and A. Wald (1943) On the Statistical Treatment of Linear Stochastic Difference Equations, Econometrika 11, 173-220.

Mehta, M. (1967) Random Matrices and Statistical Theory of Energy Levels, Academic Press, New York.

Srivastava, M. S. and C. G. Khatri (1979) Introduction to Multivariate Statistics, North Holland, New York.

Wigner, E. (1967) Random Matrices in Physics, SIAM Review 9, 1-23.

RAJ:TL:scr

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>2825 | 2. GOVT ACCESSION NO.<br>AD-A158 179 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>A LINEAR COMBINATION TEST FOR DETECTING SERIAL CORRELATION IN MULTIVARIATE SAMPLES | | 5. TYPE OF REPORT & PERIOD COVERED<br>Summary Report - no specific reporting period |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Richard A. Johnson and Thore Langeland | | 8. CONTRACT OR GRANT NUMBER(s)<br>DAAG29-80-C-0041<br>MCS-820-2205 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Mathematics Research Center, University of Wisconsin<br>610 Walnut Street<br>Madison, Wisconsin 53706 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>Work Unit Number 4 - Statistics and Probability |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>See Item 18 below. | | 12. REPORT DATE<br>June 1985 |
| | | 13. NUMBER OF PAGES<br>19 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, D. C. 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multivariate
Test of independence
Linear combination
Distribution of eigenvalues

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We propose a test for detecting serial dependence among multivariate observations. Our test statistic is the maximum absolute value of the lag 1 correlation obtainable from a linear combination of the observations. We express the statistic in terms of two eigenvalues and then obtain the asymptotic null distribution. Asymptotic power is examined for sequences of local alternatives in a multivariate normal autoregressive process. An explicit expression is obtained for the density of the limit distribution in the bivariate case. We then compare power with the likelihood ratio statistic.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

9-85

DTIC